

Generalized Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac and Acharya-Swamy Statistics and the Pólya Urn Model

Robert K. Niven^{1,2,*} and Marian Grendar^{3,†}

¹*School of Aerospace, Civil and Mechanical Engineering,
The University of New South Wales at ADFA, Northcott Drive, Canberra, ACT, 2600, Australia.*

²*Niels Bohr Institute, University of Copenhagen, Copenhagen Ø, Denmark.*

³*Department of Mathematics, Faculty of Natural Sciences,
Bel University, Tajovskeho 40, 974 01 Banská Bystrica, Slovakia.*

(Dated: 15 August 2008)

Generalized probability distributions for Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics, with unequal source probabilities q_i for each level i , are obtained by combinatorial reasoning. For equiprobable degenerate sublevels, these reduce to those given by Brillouin in 1930, more commonly given as a statistical weight for each statistic. These distributions and corresponding cross-entropy (divergence) functions are shown to be special cases of the Pólya urn model, involving neither independent nor identically distributed (“nini”) sampling. The most probable Pólya distribution contains the Acharya-Swamy intermediate statistic.

PACS numbers: 02.50.Cw, 03.65.Ta, 05.20.-y, 05.30.-d

Keywords: entropy, combinatorial, statistical mechanics, Boltzmann principle, MaxProb, quantum mechanics, boson, fermion, intermediate statistics

I. INTRODUCTION

Over 80 years ago, many physical systems were found to be governed by Bose-Einstein (BE) or Fermi-Dirac (FD) statistics [1, 2, 3, 4, 5], as distinct from classical degenerate Maxwell-Boltzmann (MB) statistics developed in the 19th century [6, 7]. The combinatorial weights of these statistics - the number of configurations (microstates) of each identifiable realization (complexion, macrostate) of the system - are widely given as [e.g. 1, 2, 3, 4, 5, 8, 9, 10, 11, 12]:

$$\mathbb{W}_{MB} = N! \prod_{i=1}^s \frac{g_i^{n_i}}{n_i!}, \quad (1)$$

$$\mathbb{W}_{BE} = \prod_{i=1}^s \frac{(g_i + n_i - 1)!}{(g_i - 1)! n_i!}, \quad (2)$$

$$\mathbb{W}_{FD} = \prod_{i=1}^s \frac{g_i!}{n_i! (g_i - n_i)!}, \quad (3)$$

where i denotes each distinguishable level (outcome or state) of the system, from a total of s levels; n_i is the number of entities in each level i ; $N = \sum_{i=1}^s n_i$ is the total number of entities; and g_i is the number of distinguishable sublevels within each level i (the degeneracy or multiplicity). In each case, the dimensionless entropy per entity H is obtained from the Boltzmann principle [6, 7]:

$$H = \frac{S_{total}}{kN} = \frac{\ln \mathbb{W}}{N}, \quad (4)$$

where S_{total} is the total dimensional (thermodynamic) entropy and k is the Boltzmann constant. Maximizing the entropy defined by (4) (“MaxEnt”), subject to the constraints on the system, is therefore equivalent to identifying the system with its most probable realization, a method of inductive reasoning which can be termed the *maximum probability* (“MaxProb”) principle [6, 7, 13, 14, 15, 16, 17, 18, 19]. Normally, (4) is subject to the asymptotic limit $N \rightarrow \infty$ (simplistically using the Stirling approximation, $\ln m! \approx m \ln m - m$, or more rigorously with Sanov’s [20] theorem), to give the entropy function for each statistic [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Recently, (4) has been applied directly to give the *exact* or *non-asymptotic* entropy function for each statistic [16, 17, 18, 19], enabling the application of statistical mechanics to systems containing small numbers of entities. The non-asymptotic BE and FD entropy functions have peculiar information-theoretic properties, with important implications for quantum mechanics [16, 17].

The above weights were generalized by Brillouin [9, also in 21, 22] to give:

$$\mathbb{P}_{MB|G} = \frac{N!}{G^N} \prod_{i=1}^s \frac{g_i^{n_i}}{n_i!}, \quad (5)$$

$$\mathbb{P}_{BE|G} = \frac{N!(G-1)!}{(G+N-1)!} \prod_{i=1}^s \frac{(g_i + n_i - 1)!}{n_i! (g_i - 1)!}, \quad (6)$$

$$\mathbb{P}_{FD|G} = \frac{N!(G-N)!}{G!} \prod_{i=1}^s \frac{g_i!}{n_i! (g_i - n_i)!}. \quad (7)$$

where \mathbb{P} denotes a correctly normalized probability distribution and $G = \sum_{i=1}^s g_i$ is the total degeneracy. These are given in many references [e.g. 23] as the probabilities $s^{-N} N! / \prod_{i=1}^s n_i!$, $\binom{s+N-1}{N}^{-1}$ and $\binom{s}{N}^{-1}$ respectively, i.e. without considering degenerate sublevels within each

*Electronic address: r.niven@adfa.edu.au

†Electronic address: marian.grendar@savba.sk

level. In contrast, the following generalized distributions were given by Kapur [24]:

$$\mathbb{P}_{BE}^{Kapur} = \prod_{i=1}^s \frac{(g_i + n_i - 1)!}{n_i!(g_i - 1)!} \frac{q_i^{n_i}}{(1 + q_i)^{g_i + n_i}}, \quad (8)$$

$$\mathbb{P}_{FD}^{Kapur} = \prod_{i=1}^s \frac{g_i!}{n_i!(g_i - n_i)!} q_i^{n_i} (1 - q_i)^{g_i - n_i}. \quad (9)$$

where q_i is the source probability¹ of an entity in level i . In the limits $q_i \ll 1$ and $g_i \gg n_i$, these imply:

$$\mathbb{P}_{MB}^{Kapur} = \prod_{i=1}^s \frac{g_i^{n_i} q_i^{n_i}}{n_i!}, \quad (10)$$

(see later discussion). The Brillouin and Kapur sets of distributions are incompatible, and their range of validity and interrelationships have not been explored.

Of great interest in recent years is the possible occurrence of “anyon” particles, intermediate between bosons and fermions. Several intermediate statistics have been proposed. A long-standing philosophical approach is to consider the allocation of indistinguishable particles to distinguishable degenerate boxes, with a maximum of d particles per box; the end-members $d \rightarrow \infty$ and $d = 1$ respectively give BE and FD statistics [28, 29, 30]. A different statistic by Haldane and Wu [31, 32] emerges from a generalized Pauli exclusion principle, involving coupled interactions between particles in occupancies $\{n_i\}$. A third approach was given as an ansatz by Acharya and Swamy [33], and later justified (approximately) by quantum symmetry [34], or by exclusion [32, 35, 36], particle kinetics [37] or fractional entanglement [38] arguments.

The aims of this work are: (i) to report the correct, generalized probability distribution for MB, BE and FD statistics containing source terms q_i , and thereby to examine the validity of (1)-(3) and (5)-(10); (ii) to examine these as special cases of the Pólya statistic, involving “neither independent nor identically distributed” (“ninin”) sampling [39, 40, 41, 42, 43, 44]; and (iii) to report the asymptotic and non-asymptotic cross-entropy (divergence) functions and minimum cross-entropy (MinXEnt) distributions for Pólya systems based on the MaxProb principle [45]. It is shown that the generalized Pólya statistic contains, and therefore justifies, the Acharya-Swamy [33] intermediate statistic.

II. DERIVATIONS

By enumeration of several simple examples: (a) a non-degenerate system; (b) a uniformly degenerate system

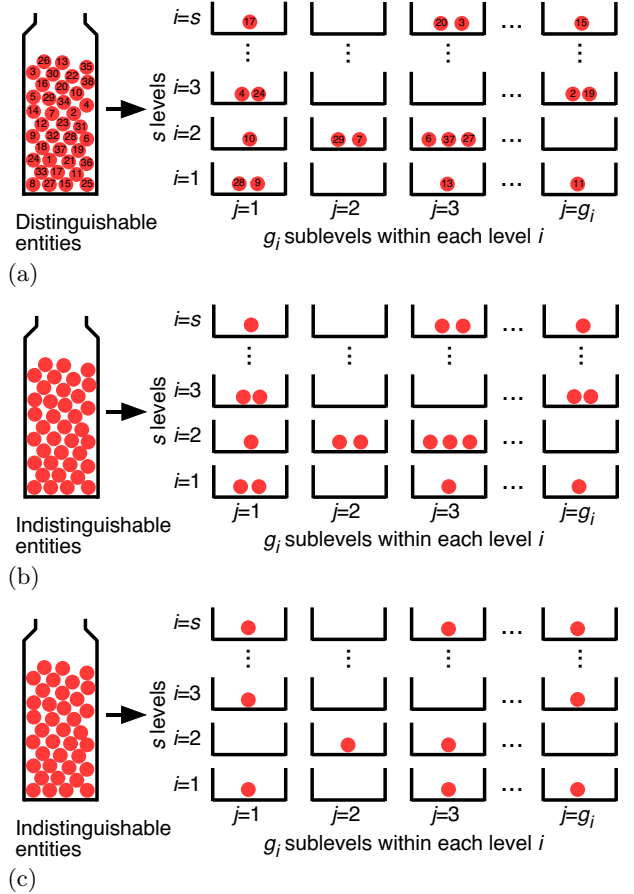


FIG. 1: Configurations of (a) Maxwell-Boltzmann, (b) Bose-Einstein and (c) Fermi-Dirac ball-in-box models, in which all N entities are fully allocated (the g_i need not be equal).

(equal g_i); (c) a non-uniformly degenerate system; and (d) systems with unequal g_i and q_i , it is readily verified that the Brillouin distributions (5)-(7) - and implicitly, the above weights (1)-(3) - are valid only for the case of equal source probabilities for each degenerate sublevel, $q_i = g_i/G$. They do not apply to equal source probabilities for each level, $q_i = s^{-1}$, or to more general forms. The Kapur distributions (8)-(10) do not satisfy normalization, $\sum_{\mathbb{P} \in \{\mathbb{P}\}} \mathbb{P} = 1$. It is therefore necessary to reexamine the probabilistic basis of each statistic.

Maxwell-Boltzmann statistics: Firstly, consider a system in which N distinguishable balls (entities) are arranged amongst s distinguishable boxes (levels), in which each level contains g_i distinguishable sublevels, as shown in Figure 1a. Each level has the source probability q_i of being filled (not necessarily equal), with $\sum_{i=1}^s q_i = 1$. We consider each *realization* of the system, containing n_i balls in each level i , $\forall i$; this in turn is a set of *configurations*, or distinguishable arrangements of balls in boxes. In counting configurations, we distinguish between rearrangements of balls between boxes, but do not (or need not) consider rearrangements of balls within each box (see Figure 1a). No account is made of the order of allo-

¹ In the MaxEnt community, q_i is termed the *prior probability*, implying a connection to Bayesian inference; the existence and form of such a connection is the subject of much debate [e.g. 25, 26, 27].

cating balls to boxes. Of course, the realizations can be chosen in many different ways; the occupancies $\{n_i\}$ are selected here for their utility in physical systems.

Clearly, the probability of an individual arrangement of n_i balls in the i th level is $q_i^{n_i}$, and so the probability of any specified configuration *amongst the levels* is $\prod_{i=1}^s q_i^{n_i}$. To obtain \mathbb{P}_{MB} , this must be multiplied by the number of ways in which the balls can be rearranged amongst the levels to give the same realization, $N! / \prod_{i=1}^s n_i!$. The result is the *non-degenerate* multinomial distribution:

$$\mathbb{P}_{MB} = N! \prod_{i=1}^s \frac{q_i^{n_i}}{n_i!}. \quad (11)$$

Another viewpoint is that if the source probability of the i th level is q_i , that of the j th sublevel of the i th level is $q_{ij} = q_i/g_i, \forall j = 1, \dots, g_i$, assuming equiprobable filling of sublevels within a level. The probability of a configuration *amongst the sublevels* - i.e. of n_{ij} balls in the j th sublevel of the i th level, $\forall \{i, j\}$ - is then $\prod_{i=1}^s \prod_{j=1}^{g_i} (q_i/g_i)^{n_{ij}} = \prod_{i=1}^s (q_i/g_i)^{\sum_{j=1}^{g_i} n_{ij}} = \prod_{i=1}^s (q_i/g_i)^{n_i}$. This must be multiplied by the number of configurations amongst sublevels, $N! \prod_{i=1}^s g_i^{n_i} / n_i!$, again giving (11). For equiprobable sublevels $q_{ij} = G^{-1}$, $q_i = g_i/G$ and (11) reduces to the Brillouin form (5). Since the total number of configurations amongst sublevels is G^N , this is proportional to the degenerate weight (1).

An important conclusion is that when considering a generalized source probability q_i unrelated to the degeneracy g_i , the latter becomes irrelevant to the probability of each configuration. We can in fact put $q_i = m_i/M$, where $m_i \in \mathbb{N}$ is a fictitious degeneracy controlled by the source probability and $M = \sum_{i=1}^s m_i$, whence:

$$\mathbb{P}_{MB} = N! \prod_{i=1}^s \frac{(m_i/M)^{n_i}}{n_i!}, \quad (12)$$

Eq. (12) has a strong connection to *urn models* of probability distributions [e.g. 23, 44, 46], as shown in Figure 2, in which balls are drawn from an urn containing M balls, made up of m_i balls of each level (color) i . In each sampling event, a ball is drawn from the urn, its color recorded, and is returned to the urn in accordance with some rule, until a sample of N balls has been recorded. In the MB case, the balls are drawn *with replacement*, hence the q_i remain constant. As shown in Figure 2, the balls are indistinguishable except for their color; the weight $N! / \prod_{i=1}^s n_i!$ now arises from the number of ways in which a given set of balls can be sampled, i.e. the number of ordered *sequences* of balls which make up each realization or *type*. Urn models thus capture many features of sets of alphabetic characters a_i drawn from an alphabet $\mathcal{A} = \{a_i\}$, considered in information theory [47, 48]. The urn model insight is of benefit to later analyses.

The above distribution can be generalized to the case of unequal sources q_{ij} for each sublevel, to give:

$$\mathbb{P}'_{MB} = N! \prod_{i=1}^s \frac{g_i^{n_i}}{n_i!} \prod_{j=1}^{g_i} q_{ij}^{n_{ij}}, \quad (13)$$

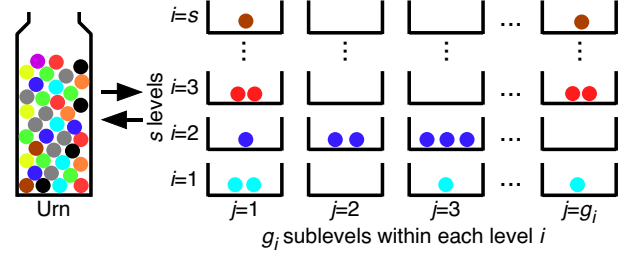


FIG. 2: An urn model, in which balls are drawn, recorded and returned to an urn, until N balls have been sampled.

with normalization $\sum_{i=1}^s \sum_{j=1}^{g_i} q_{ij} = 1$. It is unclear why such a complicated arrangement might be needed, since each sublevel could be handled as a separate level; but it does give a recipe for “coarse graining” if this is desired.

Bose-Einstein statistics: We now consider the same balls-in-boxes problem but with indistinguishable balls, as shown in Figure 1b. The standard argument is that there are $\binom{g_i+n_i-1}{n_i}$ distinguishable ways of arranging n_i indistinguishable balls in g_i boxes, $\forall i$, hence the number of configurations amongst sublevels in a realization is given by the BE weight (2). For equally probable configurations, the governing probability is obtained by dividing by the total number of configurations amongst sublevels, $\binom{G+N-1}{N}$, giving the Brillouin form (6).

We now employ the urn model analogy, to consider the generalized BE distribution:

$$\mathbb{P}_{BE} = \frac{N!(M-1)!}{(M+N-1)!} \prod_{i=1}^s \frac{(m_i+n_i-1)!}{n_i!(m_i-1)!}. \quad (14)$$

Using $q_i = m_i/M$ and parameter $\beta^- = N/M \leq 1$ gives:

$$\mathbb{P}_{BE} = \frac{N!(N/\beta^- - 1)!}{(N/\beta^- + N - 1)!} \prod_{i=1}^s \frac{(q_i N/\beta^- + n_i - 1)!}{n_i!(q_i N/\beta^- - 1)!}, \quad (15)$$

with $q_i N/\beta^- \in \mathbb{N}$. This is a variant of the *multivariate negative hypergeometric* or *Dirichlet compound multinomial distribution* [42, 49, 50], as is the Brillouin BE distribution (6) itself. This arises from a complicated scenario of *sampling without replacement*, in which two sets of balls are drawn from an urn; the conditional distribution of the second sample, given the first, is the one of interest [42, 50]. For equiprobable sublevels, $q_i = g_i/G$; taking $m_i = g_i$ and $M = G$, (14) reduces to the Brillouin form (6), and thus is proportional to the BE weight (2). For equiprobable levels $q_i = s^{-1}$, thus taking $m_i = 1$ and $M = s$, it yields the non-degenerate BE distribution $\binom{s+N-1}{N}^{-1}$. Outside these two cases, (15) gives a generalized BE distribution with unequal source probabilities q_i , which satisfies normalization and several other requirements of a governing probability distribution (see later sections).

Fermi-Dirac statistics: Here we also have indistinguishable balls, but each sublevel contains a maximum

of one ball, as shown in Figure 1c. By standard reasoning, there are $\binom{g_i}{n_i}$ distinguishable ways of arranging $n_i \in \{0, 1\}$ indistinguishable balls in g_i boxes, $\forall i$; thus the number of configurations amongst sublevels in a realization is given by the FD weight (3). For equiprobable configurations, the total number of configurations amongst sublevels is $\binom{G}{N}$, giving the Brillouin distribution (7).

Applying the urn model analogy gives the generalized FD distribution:

$$\mathbb{P}_{FD} = \frac{N!(M-N)!}{M!} \prod_{i=1}^s \frac{m_i!}{n_i!(m_i - n_i)!}, \quad (16)$$

with the condition $m_i \geq n_i$. In terms of q_i and parameter $\beta^+ = N/M \leq 1$:

$$\mathbb{P}_{FD} = \frac{N!(N/\beta^+ - N)!}{(N/\beta^+)!} \prod_{i=1}^s \frac{(q_i N/\beta^+)!}{n_i!(q_i N/\beta^+ - n_i)!}, \quad (17)$$

with $q_i N/\beta^+ \in \mathbb{N}$ and $q_i N/\beta^+ \geq n_i$. This and the Brillouin FD distribution (7) are variants of the *multivariate hypergeometric distribution* [42, 49, 50], which arises from a simple case of *sampling without replacement* [42, 49, 50]. For equiprobable sublevels, this function also reduces to the Brillouin form (7), usually expressed as the FD weight (3), whilst for equiprobable levels, it gives the non-degenerate FD distribution $\binom{s}{N}^{-1}$. Aside from these cases, (17) gives a generalized FD distribution with unequal source probabilities q_i .

III. DISCUSSION

For $a, b \in \mathbb{N}$, the following two inequalities are well known:

$$\frac{a!}{(a-b)!} \leq a^b \leq \frac{(a+b-1)!}{(a-1)!}, \quad (18)$$

with equality for $a \gg b$. Proofs are given in Appendix B. These and the weights (1)-(3) yield the inequalities [e.g. 12]:

$$\mathbb{W}_{FD} \leq \frac{\mathbb{W}_{MB}}{N!} \leq \mathbb{W}_{BE} \quad (19)$$

with equality as $g_i \gg n_i, \forall i$. If we consider the Brillouin distributions (5)-(7), one obtains:

$$\mathbb{P}_{FD|G} \gtrless \mathbb{P}_{MB|G} \gtrless \mathbb{P}_{BE|G} \quad (20)$$

The variable direction arises from the terms which converge to $\prod_{i=1}^s g_i^{n_i}/G^N$, for which opposite examples can be found (see Appendix B); equality in (20) occurs when $G \gg N$ and $g_i \gg n_i, \forall i$.

Much attention has been paid to (19), especially to the term $\mathbb{W}_{MB}/N!$, which has even been labelled a separate statistic for “corrected Boltzon” particles [12, 51]. However, comparing (19)-(20), it is evident that the $N!$

divisor in (19) arises from the fact that the MB weight is normalized differently to the BE and FD weights. Since probabilities are always normalized consistently, no such discrepancy appears in the Brillouin distributions (5)-(7), nor indeed in the generalized forms (11), (15) and (17) derived herein. In consequence, the “corrected Boltzon” statistic has no independent physical meaning, being merely an artefact of the use of weights; indeed, except in the asymptotic limit $N \rightarrow \infty$, it gives absurd results. This demonstrates an important point, that probabilistic inference using the MaxProb principle should *always* be based directly on true probabilities \mathbb{P} ; any contrary method (e.g., one based on weights \mathbb{W}) can lead to logical inconsistencies [18].

IV. THE PÓLYA DISTRIBUTION AND MAXPROB

We now consider the Pólya urn model (Figure 2), in which a ball is sampled from an urn, recorded, and returned to the urn; in addition, $c \in \mathbb{Z}$ balls of the same color (level) are also added to the urn. The probability of drawing the realization $\{n_i\}$ from the urn after N draws is [39, 40, 41, 42, 43, 44]:

$$\mathbb{P}_{Polya} = \frac{N!}{\prod_{i=1}^s n_i!} \frac{\prod_{i=1}^s m_i(m_i + c) \cdots (m_i + (n_i - 1)c)}{M(M+c) \cdots (M + (N-1)c)}, \quad (21)$$

The Pólya urn thus describes a simple, analytic case of “neither independent nor identically distributed” (“niniid”) sampling, of which the bivariate form is widely used in studies of biological contagion [e.g. 43].

The Pólya model (21) can be written in the following forms, valid respectively for $c = 0$, $c > 0$ and $c < 0$:

$$\mathbb{P}_{Polya}^{c=0} = \frac{N!}{M^N} \prod_{i=1}^s \frac{m_i^{n_i}}{n_i!}, \quad (22)$$

$$\mathbb{P}_{Polya}^{c>0} = \frac{N!}{\prod_{i=1}^s n_i!} \frac{\Gamma(\frac{M}{c})}{\Gamma(\frac{M}{c} + N)} \prod_{i=1}^s \frac{\Gamma(\frac{m_i}{c} + n_i)}{\Gamma(\frac{m_i}{c})}, \quad (23)$$

$$\mathbb{P}_{Polya}^{c<0} = \frac{N!}{\prod_{i=1}^s n_i!} \frac{\Gamma(1 - \frac{M}{c} - N)}{\Gamma(1 - \frac{M}{c})} \prod_{i=1}^s \frac{\Gamma(1 - \frac{m_i}{c})}{\Gamma(1 - \frac{m_i}{c} - n_i)}. \quad (24)$$

These distributions encompass the MB, BE and FD governing distributions (12), (14) and (16). Simplified forms of (21) and (23)-(24) were given by [36, 37]. In light of the Pólya model, we can set aside the complicated urn model for the multivariate negative hypergeometric distribution (the generalized BE statistic (14)), to offer the following explanations for each statistic:

- MB statistic: Pólya urn with $c=0$ (sampling with replacement);
- BE statistic: Pólya urn with $c=1$ (sampling with double replacement);

- FD statistic: Pólya urn with $c=-1$ (sampling without replacement);

It is curious that the last two representations are equivalent to the allocation of indistinguishable balls to distinguishable boxes (with a further restriction in the FD case). Whilst the full implications of this insight are not clear, it does provide an alternative philosophical foundation for the construction of quantum statistics.

The Pólya distributions (23)-(24) can be further analyzed using an extension of Boltzmann's principle (4):

$$D = -\frac{\ln \mathbb{P}}{N}, \quad (25)$$

where D is the generalized cross-entropy (directed divergence or negative relative entropy) function defined by the MaxProb principle. Applying (25) in the asymptotic limit $N \rightarrow \infty$ to (22)-(24), using $p_i = n_i/N$, $q_i = m_i/M$ and $\beta = N/M$, gives:

$$-D_{Polya} = \sum_{i=1}^s -p_i \ln \frac{p_i}{q_i} \quad (26)$$

$$-D_{Polya} = \sum_{i=1}^s \left\{ -p_i \ln p_i + p_i \left(\frac{1}{\beta c} \right) \ln \left(\frac{1}{\beta c} \right) - p_i \left(\frac{1}{\beta c} + 1 \right) \ln \left(\frac{1}{\beta c} + 1 \right) \right. \quad (27)$$

$$\left. + \left(\frac{q_i}{\beta c} + p_i \right) \ln \left(\frac{q_i}{\beta c} + p_i \right) - \left(\frac{q_i}{\beta c} \right) \ln \left(\frac{q_i}{\beta c} \right) \right\},$$

$$-D_{Polya} = \sum_{i=1}^s \left\{ -p_i \ln p_i - p_i \left(-\frac{1}{\beta c} \right) \ln \left(-\frac{1}{\beta c} \right) + p_i \left(-\frac{1}{\beta c} - 1 \right) \ln \left(-\frac{1}{\beta c} - 1 \right) - \left(-\frac{q_i}{\beta c} - p_i \right) \ln \left(-\frac{q_i}{\beta c} - p_i \right) + \left(-\frac{q_i}{\beta c} \right) \ln \left(-\frac{q_i}{\beta c} \right) \right\}. \quad (28)$$

with restriction $|c| < q_i/\beta p_i, \forall i$ (whence $|c| < m_i/n_i, \forall i$) for $c < 0$. Note (26) is the Kullback-Liebler cross-entropy [52]. Proof of these results based on Sanov's theorem is given elsewhere [45]; a more simplistic calculation may be conducted by applying Stirling's approximation to all factorials and gamma functions. Both forms of the Pólya cross-entropy (27)-(28) have been shown to satisfy many important information-theoretic properties, including non-negativity, lower semicontinuity, convexity, partition inequality and the conditional limit theorem [45].

Applying the MaxProb principle, the relevant cross-entropy (26), (27) or (28) can be minimized (MinXEnt) subject to the following constraints:

$$\sum_{i=1}^s p_i = 1. \quad (29)$$

$$\sum_{i=1}^s p_i f_{ri} = \langle f_r \rangle, \quad r = 1, \dots, R, \quad (30)$$

where f_{ri} is the value of the function f_r in the i th level and $\langle f_r \rangle$ is its mathematical expectation. Expremization by the Lagrangian method gives the inferred "most probable" distribution for each statistic:

$$p_{Polya,i}^* = q_i e^{-\lambda_0^{c=0} - \sum_{r=1}^R \lambda_r f_{ri}}, \quad (31)$$

$$p_{Polya,i}^* = \frac{q_i/\beta c}{e^{\lambda_0^{c>0} + \sum_{r=1}^R \lambda_r f_{ri}} - 1}, \quad (32)$$

$$p_{Polya,i}^* = -\frac{q_i/\beta c}{e^{\lambda_0^{c<0} + \sum_{r=1}^R \lambda_r f_{ri}} + 1}, \quad (33)$$

where $\lambda_r, r = 0, \dots, R$ are Lagrangian multipliers, with constants absorbed into the λ_0 terms. For equiprobable sublevels $q_i = g_i/G$, sample size $\beta = N/G$ and the above values of c , the cross-entropy functions (26)-(28) and distributions (31)-(33) reduce to those given by Brillouin [8, 9]; these in turn simplify - up to constant factors - to the commonly used entropy functions and distributions obtained from the weights² (1)-(3) [1, 2, 3, 4, 5, 8, 9, 10, 11, 12].

For $N \ll \infty$, the non-asymptotic cross-entropy functions and distributions for the above systems can also be derived using Boltzmann's principle (25), as reported in Appendix A. These encompass the non-asymptotic MB, BE and FD entropy functions and distributions reported previously [16, 17].

To summarize, generalized MB, BE and FD statistics, containing unequal source probabilities for each level, can be obtained using the combinatorial definition of cross-entropy (25) in conjunction with probability distributions generated using urn models. All such cases emerge as special instances of the Pólya urn model, involving a simple scenario of urn modification during sampling. Of the systems examined, only the MB system in the asymptotic limit yields the Kullback-Leibler cross-entropy (26) [52]; in consequence, the latter is *not* universal in application, and *cannot* be used to infer the most probable realization of BE, FD or Pólya systems except in special limiting cases (nor even of the MB system away from the asymptotic limit $N \rightarrow \infty$). This conclusion contradicts the dominant viewpoint in information theory, in which the Shannon entropy [47] and Kullback-Liebler cross-entropy [52] are considered to be the paramount functions for inductive reasoning ("statistical inference"), based on their axiomatic and information-theoretic definitions [c.f. 53, 54]. This conclusion is, however, a necessary consequence of Boltzmann's principle.

Eqs. (31)-(33) can be amalgamated as:

$$p_{Polya,i}^* = \frac{q_i}{e^{\lambda_0' + \sum_{r=1}^R \lambda_r f_{ri}} - \beta c}, \quad (34)$$

² Physicists typically consider constraints on normalization (29) and energy $f_{1i} = E_i$, $\langle f_1 \rangle = \langle E \rangle$ in (30), giving multipliers $\lambda_1 = 1/kT$ and $\lambda_0 = -\mu/kT$, where k is the Boltzmann constant, T is absolute temperature and μ can be viewed as a chemical potential.

This can be compared to the ansatz presented by Acharya and Swamy [33], in the present notation:

$$p_{AS,i}^* = \frac{g_i/N}{e^{\lambda'_0 + \sum_{r=1}^R \lambda_r f_{ri}} - a}, \quad (35)$$

in which $a = 0, 1$ and -1 respectively give MB, BE and FD statistics, and $-1 \leq a \leq 1$ their intermediate interpolation. The Pólya representation (34) thus contains the Acharya-Swamy intermediate statistic (35) as a special case. It therefore justifies this statistic using a “ninid” urn sampling scheme, an explanation rather different to that of previous studies [32, 34, 35, 36, 37, 38].

V. CONCLUSIONS

Generalized probability distributions for MB, BE and FD statistics, with unequal source probabilities q_i for each level i , are derived by extension of the distributions given by Brillouin [8, 9], using an urn model analogy. These are shown to be special instances of the Pólya urn model, involving a simple scenario of urn modification during sampling (“ninid” sampling). The resulting cross-entropy functions and MinXEnt distributions are derived by Boltzmann’s principle (the MaxProb method). The general form of the Pólya distribution contains the Acharya-Swamy [33] intermediate statistic, as an exact result without approximation.

We also show that the “corrected Boltzon” statistic based on $\mathbb{W}_{MB}/N!$ has no independent meaning, but is an artefact of the use of weights instead of probabilities in Boltzmann’s principle. Furthermore, since the Pólya cross-entropy function in general differs from the Kullback-Liebler cross-entropy [52], the latter does not give the most probable realization of BE, FD or Pólya systems except in special limiting cases.

As a final comment, the statistic of Haldane and Wu [31, 32] considers intermediacy as a consequence of changes in occupancies $\{n_i\}$ due to interactions. The Acharya-Swamy-Pólya model considers intermediacy due to changes in the source probabilities $\{q_i\}$ - hence in the degeneracies $\{g_i\}$ - during sampling. Either approach is justifiable on physical grounds; their common features and possible synthesis deserve further investigation.

Acknowledgments

The first author was supported by a Marie Curie Incoming International Fellowship project 039729, awarded by the European Commission under Framework Programme 6. The second author was supported by VEGA grant 1/3016/06 and Australian Research Council grant no. DP0210999; the hospitality of the School of Computer Science and Engineering, The University of New South Wales, Australia (with special thanks to Arthur Ramer) and of the Niels Bohr Institute, University of Copenhagen, Denmark, are gratefully acknowledged.

VI. APPENDIX A: NON-ASYMPTOTIC DISTRIBUTIONS

The non-asymptotic forms of the Pólya cross-entropy functions can be obtained directly from Boltzmann’s principle (25), without taking the asymptotic limit $N \rightarrow \infty$. Such statistics are required for systems of small numbers of non-interacting entities [16, 17, 18, 19]. From (25) and the governing distributions (11), (15) and (17), the functions are:

$$-D_{Polya}^{(N)} = \frac{1}{N} \sum_{i=1}^s \{p_i \ln N! + p_i N \ln q_i - \ln(p_i N)!\} \quad (36)$$

$$-D_{Polya}^{(N)} = \frac{1}{N} \sum_{i=1}^s \left\{ p_i \ln \frac{N! \Gamma(N/\beta c)}{\Gamma(N/\beta c + N)} + \ln \frac{\Gamma(q_i N/\beta c + p_i N)}{(p_i N)! \Gamma(q_i N/\beta c)} \right\} \quad (37)$$

$$-D_{Polya}^{(N)} = \frac{1}{N} \sum_{i=1}^s \left\{ p_i \ln \frac{N! \Gamma(1 - N/\beta c - N)}{\Gamma(1 - N/\beta c)} + \ln \frac{\Gamma(1 - q_i N/\beta c)}{(p_i N)! \Gamma(1 - q_i N/\beta c - p_i N)} \right\} \quad (38)$$

in which constant terms are brought inside the sum using (29). Applying MinXEnt subject to (29)-(30) gives the inferred “most probable” distribution for each statistic:

$$p_{Polya,i}^{\#} = \frac{1}{N} \Lambda^{-1} \left\{ \frac{1}{N} \ln N! + \ln q_i - \lambda_0^{c=0} - \sum_{r=1}^R \lambda_r f_{ri} \right\} \quad (39)$$

$$p_{Polya,i}^{\#} = \frac{1}{N} \Lambda^{-1} \left\{ \frac{1}{N} \ln \frac{N! \Gamma(N/\beta c)}{\Gamma(N/\beta c + N)} + \Lambda(q_i N/\beta c + p_{Polya,i}^{\#} N - 1) - \lambda_0^{c>0} - \sum_{r=1}^R \lambda_r f_{ri} \right\} \quad (40)$$

$$p_{Polya,i}^{\#} = \frac{1}{N} \Lambda^{-1} \left\{ \frac{1}{N} \ln \frac{N! \Gamma(1 - N/\beta c - N)}{\Gamma(1 - N/\beta c)} - \Lambda(-q_i N/\beta c - p_{Polya,i}^{\#} N) - \lambda_0^{c<0} - \sum_{r=1}^R \lambda_r f_{ri} \right\} \quad (41)$$

where $\lambda_r, r = 0, \dots, R$ again are Lagrangian multipliers, $\Lambda(x) = \psi(x+1) = y$ is a shifted digamma function, defined for convenience, and $\Lambda^{-1}(y) = \psi^{-1}(y) - 1$ is its inverse. None of these functions allow explicit factorization of a partition function $Z = e^{\lambda_0}$; the last two forms are also implicit in $p_i^{\#}$.

VII. APPENDIX B: INEQUALITIES

For $b > 0$ the terms in (18) can be written:

$$\frac{a!}{(a-b)!} = a(a-1)\dots(a-b+1) \quad (42)$$

$$\frac{(a+b-1)!}{(a-1)!} = (a+b-1)(a+b-2)\dots a \quad (43)$$

As $a/b \rightarrow \infty$, each factor in (42)-(43) will approach a respectively from below or above. Both forms thus converge to a^b in accordance with (18); (19) follows. \square

From (42)-(43), both $\mathbb{P}_{BE|G}$ and $\mathbb{P}_{FD|G}$ converge to $\mathbb{P}_{MB|G}$. However, the sign of the inequalities can vary. This is illustrated by the set of realizations for the Brillouin case $N = 4$, $s = 3$ and $g_i = 3, \forall i$, for which the probabilities (5)-(7) are listed in Table I (some realizations are excluded from the FD set). Depending on the realizations, the order can be $\mathbb{P}_{FD|G} < \mathbb{P}_{MB|G} < \mathbb{P}_{BE|G}$, $\mathbb{P}_{FD|G} > \mathbb{P}_{MB|G} > \mathbb{P}_{BE|G}$ or $\mathbb{P}_{FD|G} < \mathbb{P}_{MB|G} > \mathbb{P}_{BE|G}$. In this and many other examples calculated, we did not observe $\mathbb{P}_{FD|G} > \mathbb{P}_{MB|G} < \mathbb{P}_{BE|G}$, although this result is not proven.

Realization [n_1, n_2, n_3]	\mathbb{P}_{FD}	\geq \leq	\mathbb{P}_{MB}	\geq \leq	\mathbb{P}_{BE}
[1, 1, 2]	0.21429	>	0.14815	>	0.10909
[1, 2, 1]	0.21429	>	0.14815	>	0.10909
[2, 1, 1]	0.21429	>	0.14815	>	0.10909
[2, 2, 0]	0.07143	<	0.07407	>	0.07273
[2, 0, 2]	0.07143	<	0.07407	>	0.07273
[0, 2, 2]	0.07143	<	0.07407	>	0.07273
[1, 3, 0]	0.02381	<	0.04938	<	0.06061
[1, 0, 3]	0.02381	<	0.04938	<	0.06061
[3, 1, 0]	0.02381	<	0.04938	<	0.06061
[3, 0, 1]	0.02381	<	0.04938	<	0.06061
[0, 1, 3]	0.02381	<	0.04938	<	0.06061
[0, 3, 1]	0.02381	<	0.04938	<	0.06061
[4, 0, 0]	NA		0.01235	<	0.03030
[0, 4, 0]	NA		0.01235	<	0.03030
[0, 0, 4]	NA		0.01235	<	0.03030

TABLE I: Probabilities of each realizations for $N = 4$, $s = 3$ and $g = [3, 3, 3]$ (NA=not applicable).

-
- [1] S.N. Bose, Z. Phys. 26 (1924) 178.
 - [2] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl (1924) 261.
 - [3] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl (1925) 3.
 - [4] E. Fermi, Z. Phys. 36 (1926) 902.
 - [5] P.A.M. Dirac, Proc. Roy. Soc. 112 (1926) 661.
 - [6] L. Boltzmann, Wien. Ber. 76 (1877) 373, English transl., J. Le Roux (2002), <http://www.essi.fr/~leroux/>.
 - [7] M. Planck, Annalen der Physik 4 (1901) 553.
 - [8] L. Brillouin, Annales de Physique 7 (1927) 315.
 - [9] L. Brillouin, *Les Statistiques Quantiques et Leurs Applications*, Les Presses Universitaires de France, Paris, 1930.
 - [10] R.C. Tolman, *The Principles of Statistical Mechanics*, Oxford Univ. Press, London, 1938, chap. X.
 - [11] L. Brillouin, J. Appl. Phys. 22(3) (1951) 338.
 - [12] N. Davidson, *Statistical Mechanics*, McGraw-Hill, NY, 1962, chap. 6.
 - [13] I. Vincze, Progress in Statistics 2 (1974) 869.
 - [14] M. Grendár, Jr. M. Grendár, in A. Mohammad-Djafari (ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2000)*, AIP Conf. Proc., 2001, 83.
 - [15] R. K. Niven, *cond-mat/0512017*, 2005-2007.
 - [16] R. K. Niven, Phys. Lett. A, 342(4) (2005) 286.
 - [17] R. K. Niven, Physica A, 365(1) (2006) 142.
 - [18] R.K. Niven, in Knuth, K.H., Caticha, A., Center, J.L., Giffon, A., Rodriguez, C.C. (eds), *MaxEnt 2007*, Saratoga Springs, NY, 8-13 July 2007, AIP Conf. Proc. 954, 133.
 - [19] R.K. Niven, *cond-mat/0807.4160*, 2008.
 - [20] I. N. Sanov, Mat. Sbornik, 42 (1957) 1144 (Russian).
 - [21] R. Fortet, *Elements of Probability Theory*, Gordon and Breach Science Publ., London, 1977.
 - [22] C.B. Read, in S. Kotz, N.L. Johnson, *Encyclopedia of Statistical Sciences*, vol. 3, John Wiley, NY, 1983, 63-66.
 - [23] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed., John Wiley, NY.
 - [24] J.N. Kapur, Bull. Math. Assoc. India 21 (1989) 39 (note: contains typographical errors).
 - [25] K.H. Knuth, Neurocomputing 67 (2005) 245.
 - [26] A. Caticha, in Knuth, K.H., Caticha, A., Center, J.L., Giffon, A., Rodriguez, C.C. (eds), *MaxEnt 2007*, Saratoga Springs, NY, 8-13 July 2007, AIP Conf. Proc. 954, 11.
 - [27] M. Grendar, *math.ST/0804.3926*, 2008.
 - [28] G. Gentile, Nuovo Cim. 17 (1940) 493.
 - [29] A.M. Guénault, D.K.C. MacDonald, Molec. Phys. 5(5) (1962) 525.
 - [30] J.N. Kapur, H.K. Ksevan, *Entropy Optimization Principles with Applications*, Academic Press, Inc., Boston, MA, 1992.
 - [31] F.D.M. Haldane, Phys. Rev. Lett. 67(8) (1991) 937.
 - [32] Y.-S. Wu, Phys. Rev. Lett. 73(7) (1994) 922.
 - [33] R. Acharya, P.N. Swamy, J. Phys. A: Math. Gen. 27 (1994) 7247.
 - [34] R. Acharya, P.N. Swamy, Physica A 340 (2004) 427.
 - [35] A.P. Polychronakos, Phys. Lett. B, 365 (1996) 202.
 - [36] A.V. Ilinskaya, K.N. Ilinsky, J.M.F. Gunn, Nucl. Phys. B, 458 (1996) 562.
 - [37] G. Kaniadakis, A. Lavagno, P. Quarati, Nucl. Phys. B 466 (1996) 527.
 - [38] T. Zhou, Solid State Comm. 115 (2000) 185.
 - [39] F. Eggenberger, G. Pólya, Über die Statistik verketteter Vorgänge Z. Angew. Math. Mech., 1 (1923) 279-289.
 - [40] G. Pólya, Ann. Inst. Henri Poincaré 1 (1931) 117.
 - [41] H. S. Steyn, Proc. Koninklijke Nederlandse Akademie van Wetenschappen, Ser. A, 54 (1951) 23-30.
 - [42] G.P. Patil, M.T. Boswell, M.V. Ratnaparkhi, J.J.J. Roux, Dictionary and Classified Bibliography of Statis-

- tical Distributions in Scientific Work, Int. Coop. Publ. House, Fairland, Maryland, 1984, 3: chap. 10.
- [43] G.P. Patil, M.V. Ratnaparkhi, Pólya Distribution, Multivariate, in S. Kotz, N.L. Johnson, *Encyclopedia of Statistical Sciences*, 1986, 7: 59.
 - [44] N. L. Johnson, S. Kotz, N. Balakrishnan, Discrete Multivariate Distributions. New York: Wiley, 1997.
 - [45] M. Grendar, R.K. Niven, *cond-mat/0612697*, 2006.
 - [46] S. Berg, Urn Models, in S. Kotz, N.L. Johnson, *Encyclopedia of Statistical Sciences*, 1988, 9: 424.
 - [47] C.E. Shannon, Bell Sys. Tech. J. 27 (1948) 379; 623.
 - [48] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd ed., John Wiley, NY, 2006.
 - [49] N.L. Johnson, S. Kotz, Discrete Distributions, Houghton Mifflin Co, Boston, chap. 11.
 - [50] D.R. Jensen, Multivariate Distributions, in S. Kotz, N.L. Johnson, *Encyclopedia of Statistical Sciences*, 1985, 6: 5200.
 - [51] R.H. Fowler, Statistical Mechanics, 2nd ed., Cambridge U.P., 1936.
 - [52] S. Kullback, R.A. Leibler, Annals Math. Stat. 22 (1951) 79.
 - [53] E.T. Jaynes, Phys. Rev. 106 (1957) 620.
 - [54] E.T. Jaynes, (G.L. Bretthorst, ed.) Probability Theory: The Logic of Science, Cambridge U.P., Cambridge, 2003.